

Understanding the Role of Trust in Human-Autonomy Teaming

Nathan J. McNeese¹, Mustafa Demir², Erin Chiou², Nancy Cooke², Giovanni Yanikian²

¹Human-Centered Computing, Clemson University
mcneese@clemson.edu

²Human Systems Engineering, Arizona State University
{mdemir, erin.chiou, Nancy.Cooke, gyanikia}@asu.edu

Abstract

This study aims to better understand trust in human-autonomy teams, finding that trust is related to team performance. A wizard of oz methodology was used in an experiment to simulate an autonomous agent as a team member in a remotely piloted aircraft system environment. Specific focuses of the study were team performance and team social behaviors (specifically trust) of human-autonomy teams. Results indicate 1) that there are lower levels of trust in the autonomous agent in low performing teams than both medium and high performing teams, 2) there is a loss of trust in the autonomous agent across low, medium, and high performing teams over time, and 3) that in addition to the human team members indicating low levels of trust in the autonomous agent, both low and medium performing teams also indicated lower levels of trust in their human team members.

1. Introduction

In recent years, human-machine interaction has been studied in a multitude of research communities ranging from human factors, information science, and computer science, and from multiple different focuses (communication, awareness, trust, etc.). Given the increasing pace of technological developments in recent years, our understanding of human-machine interaction is still relatively limited (in comparison to other domains). However, there is little doubt of its importance and longevity, moving forward.

As the concept of human-machine interaction has grown, it has branched out into related areas. Most notably, human-machine *teaming* has become an area of interest. Broadly speaking, human-machine teaming is when a human and a machine have interdependent roles, that require interaction, to achieve a common goal. The machine can consist of many different technologies, such as a hazard warning system, a virtual agent (e.g., Siri), a decision

support system (e.g., IBM's Watson), or an embodied robot with artificial intelligence (e.g., Baxter).

Although much discussion has surrounded the usage of "autonomy" as a descriptor for any type of machine, no matter how capable it is [1]–[4], we follow the practical convention of [5] and [6] and use the word to refer to machines that have the ability and freedom to make decisions and take actions on its own. Traditionally, human-machine teaming is either studied from the perspective of human-*automation* teaming or human-*autonomy* teaming. The differences between these two concepts lie within the freedom that is given to the technological "machine". Human-automation teaming is dependent on technology grounded in principles of automation, where technology will take over and perform tasks but only if the human controls and directs it to do so. Whereas, with human-autonomy teaming, the technology is autonomous, meaning that the technology has the ability and freedom to make choices on its own (i.e., the human is not in direct or supervisory control of the machine) [7]. From a teaming perspective, human-autonomy teaming is more representative of human-human teaming because its team members each have their own expertise and the authority to take action. They interact as peers, rather than as supervisors and their subordinates [8]. This difference in relational structure has important implications for the way we think about and conduct research in human-machine interaction.

Yet, very little is empirically known about human-autonomy teaming due to prior limited machine capabilities. In more recent years, the advent of advanced cognitive modeling techniques [9] and artificial intelligence [10] has brought to light the possibility of machines serving as peer-like team mates, and paved the way for studying human-autonomy teams.

More research on human-autonomy teaming is needed. We are at an inflection point where there is significant need to validate, translate, and possibly transfer the assumptions and findings of multiple

canonical human-human teaming characteristics, and human-machine interaction concepts, to the new paradigm of human-autonomy teaming. We need to know how human-human teams are similar and different from human-autonomy teams. There is often an assumption that these two types of teams will be similar, and that human-human teaming may be the ideal model for human-autonomy teaming, but we simply do not know beyond the findings of a few early studies. These assumptions need to be tested and verified.

One such human-human characteristic that needs to be further investigated in human-autonomy teams is that of trust, which is found to be fundamental to human-human teaming [31]. Trust is one of many fundamental characteristics of human-human teaming but is often identified as a major concept, and one that has been studied in many different ways [46], [47]. Specifically, trust has shown to impact team performance, indicating that teams with higher levels of trust perform better [48]. The concept of trust has substantial history within the domain of human-machine interaction [11]–[14] with only preliminary work in human-autonomy teaming [15]. Compounding and furthering the significance of the lack of work on trust in human-autonomy teaming is that autonomy has increased degrees of freedom to either develop or diminish trust both at an individual and team level. In this paper, we present a study focused on better understanding trust in human-autonomy teaming with a specific focus on how trust throughout the team is impacted, and how human team members trust a synthetic virtual agent operating in the role of a fully capable and autonomous team member in a remotely piloted aircraft system (RPAS) task and simulation setting.

1.1 An Overview of The Research Study & Research Objectives

This study focuses on the concept of trust in human-autonomy teaming within the context of the development, integration, and validation of a computational cognitive model acting as a teammate on an otherwise all-human team. The team consists of three heterogeneous roles working to complete a command-and-control task (i.e., team control of an RPAS).

We brought teams into the lab and used a wizard-of-oz methodology to simulate an autonomous agent as a team member. This methodology allows an experimenter to play the role of an autonomous teammate (communicating and coordinating information in real time to human team members) without having to actually program a synthetic teammate, and without the other team

members realizing that it is actually a human controlling the synthetic team member.

Our main research objective of this study focuses on the relationship of team performance and team social behaviors (specifically those related to trust) in human-autonomy teams over time. More specifically, we are interested in if high or low performing teams trust their autonomous teammate more or less. We are also directly interested in how trust changes or adapts over time in high and low performing teams (high and low). These objectives will help to inform the effect that trust has on team performance in human autonomy teams. The paper will proceed with a background pertaining to trust and human-autonomy teaming, an overview of methods used in this empirical effort, results, and a discussion highlighting important findings and suggestions for moving forward.

2. Trust and Human-Autonomy Teaming

2.1. Human-Autonomy Teaming

Autonomous machines that can function in roles typically occupied by human team members will likely increase in the years to come, particularly as work environments become more virtual. Such machines should be able to maintain appropriate trust over time leading to higher team performance. If goals are not met, and uncertainty is high within the human-autonomy team, then research has shown teams will perceive higher workload and will be less able to complete multiple tasks [16]. Effective human-autonomy teaming must consider several issues, such as: creating an autonomous agent that can efficiently work with humans, modeling the team interactions, and modeling human cognition to incorporate in autonomy design [17], [18]. Therefore, human-autonomy teaming research is typically grounded in the fields of computer intelligence [19], cognitive science [20], [21], and team cognition [22] to create the necessary conditions and understanding for effective human-autonomy teaming.

Until recently, there has been a lack of empirical human-autonomy teaming research. McNeese and colleagues [7] recently published a study looking at multiple team characteristics in a human-autonomy team. The human-autonomy teaming took place within a simulated RPAS and the humans were able to chat with an autonomous agent, in real time, using restricted natural language [23]. Findings from this study highlight that human-autonomy teams performed as well as human-human teams in the same simulation, but human-autonomy teams were deficient in aspects of team-level communication and

coordination. Complimentary research using this same simulation for the purposes of human-autonomy teaming has also outlined the importance of situational awareness [24] and team synchrony [25].

In addition, other work has developed a naturalistic decision-making based cognitive agent that helps people make decisions by improving situational awareness but increases cognitive load [26]. Other research has utilized a wizard-of-oz methodology to simulate an autonomous agent as a team player [27]. Research that tested human-autonomy teaming in a joint resource management and scheduling task also found that different cooperative strategies of a machine agent's behaviors [28] and social exchange structures [29] can impact human teaming behavior and overall team performance.

2.2 Trust in Human-Autonomy Teaming

Trust, as a psychological state, is an important construct needed to develop confidence and productive interactions among team members. Trust in human-human teams has been defined as “the extent to which a person is confident in, and willing to act on the basis of, the words, actions, and decisions of another” [30].

As previously noted, an autonomous agent used in human-autonomy teaming can come in many different forms (e.g., software agent, embodied robots) and will need to participate in developing a level of trust within the team. The level of team trust could be a critical element in determining how teams will perform in situations that are suboptimal.

Trust is a multidimensional construct that has been investigated in different areas of research, interpersonal relationships [30], in teams [31], and with automation [11]. Specific to the study presented in this paper, we adopt [30]'s definition of trust for the context of human-autonomy teaming. Specific to this study, our focus is during situations under degraded conditions, when different levels of trust are likely to have the most observable impact on team performance.

Organizing trust from a team perspective should be considered by affording three characteristics of trustworthiness [32]: 1) ability (e.g., what action the autonomous agent can perform and the skill associated), 2) integrity (e.g., agent-based policy), and 3) benevolence (e.g., agent intentions). These characteristics allow autonomous agents to be perceived and recognizable as trustworthy for human team members. Ability is the agent's skills, competencies, and characteristics that enable individual team members to have influence within a

specific domain. Integrity, a set of policies, will have a set degree of acceptability that autonomous agents will need to adhere to. In agent and distributed computing model contexts, the policy can be defined as “an enforceable, well-specified constraint on the performance of a machine-executable action by a subject in a given situation” [33]. Benevolence is the motivation that should align to some extent to the team goal, and includes the willingness to sacrifice individual goals for the team goal [28].

Many factors related to trust in automation and autonomy have been investigated. Olesen and colleagues [34] outline that there are multiple variables that can impact a human's level of trust in an autonomous teammate: 1) *human* influences (such as, individual differences in personality, experience, and culture), 2) *machine* influences (such as, robotic platform, robot performance in relation to automation, failure rates, and false alarms), and 3) *environmental* influences (such as, task type, and operational environment). Two specific areas that have garnered a substantial amount of attention in the trust and HAT domain are that of *anthropomorphism* and *transparency*.

de Visser and colleagues [35] found that anthropomorphic cues reduced initial expectations of automation to execute flawlessly, and when trust was violated there was a higher chance that trust would be repaired. This suggests that qualities such as perceived agency, intentionality, physical presence, and biological motion can elicit anthropomorphic perceptions of autonomous agents. Therefore, anthropomorphism could be one additional way to investigate a team's ability to trust an autonomous agent. Recent work examining how different levels of intelligent agent transparency impact trust in human-agent teaming found that trust increased as a function of transparency level, meaning that the more transparent the agent was, the more trust humans had in it [36]. Finally, scholars have highlighted the importance of communicating intent to engender trust in human-autonomy teaming [37], and understanding how to repair trust in human-autonomy teaming [38].

3. Methods

3.1 Participants & Experimental Context

This study was conducted in the Cognitive Engineering Research on Team Task (CERTT) RPAS-Synthetic Task Environment (STE) [27]. The RPAS-STE is based on the United States Air Force Predator RPAS ground control station. The RPAS-STE task requires three different, interdependent teammates within the RPAS team; each with a unique role relevant to the team's objective of efficiently

taking good photos of target waypoints. Further, the CERTT RPAS-STE is dynamic and taking good photographs of designated waypoints requires information to be shared among teammates in a timely manner.

A total of 44 participants from a large southwestern university community were recruited and participated in the study. Two participants per team were formed (22 teams) to fulfill photographer and navigator roles, and the pilot position was filled by a well-trained experimenter who mimicked a synthetic agent in terms of communication and coordination (Figure 1). Participants were informed that the pilot was a synthetic agent, but not that it was a trained experimenter.

Specific information about each role follows. 1) navigator – provides a dynamic flight plan and sends the information to the pilot about the targeted altitude and the airspeed of the current target waypoint; 2) pilot – controls the RPA’s heading, altitude, and airspeed, and negotiates with the photographer about the targeted speed and altitude restrictions for the current target waypoint); and 3) photographer – monitors sensor equipment, negotiates with the pilot, and takes photographs of target waypoints, and sends feedback regarding whether the team has a good photo or not.

Participation required normal or corrected-to-normal vision and fluency in English. Participants ranged in age from 18 to 36 ($M_{age} = 23$, $SD_{age} = 3.90$), with 21 participants self-reporting as male, and 23 as female, and were either graduate or undergraduate students. Each team participated in two \approx seven-hour sessions (with one or two-week lapsing between the two sessions), and each individual was compensated for participation by payment of \$10 per hour. Two separate sessions were completed to empirically evaluate if trust changed over time and to better inform how autonomy should change or interact with humans over an extended time period.

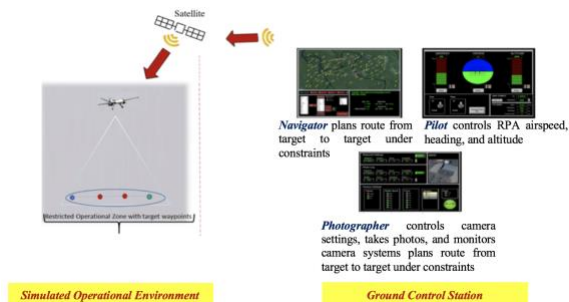


Figure 1. CERTT RPAS-STE team roles and task (the red dashed line indicates two different

environments: the simulated operational environment and the ground control station; modified from [24]).

3.2 Experimental Task and Procedure

The task was carried out over a series of 10 missions (see Table 1), wherein all interactions took place via a text-based communications system. A single RPAS-STE mission consists of 11-12 targets and lasts a maximum of 40 minutes. After signing consent forms, the participants were randomly assigned to their roles and started their role-specific training. The navigator and the photographer sat in one room (but separated via partitions, and no face-to-face contact), and the pilot sat another room.

Table 1. Experimental Sessions and Task Duration

Session-I (Total Session with breaks \approx 6 hours)	Session-II (Total Session with breaks \approx 7 hours)
1) Consent forms (15 min)	1) Mission 5 (40 min),
2) PowerPoint (30 min) and hands on training (30 min)	2) NASA TLX I (15 min)
3) Mission 1 (40 min)	3) Mission 6 (40 min),
4) NASA TLX I (15 min)	4) Mission 7 (40 min),
5) Missions 2 (40 min)	5) Mission 8 (40 min),
6) Mission 3 (40 min),	6) Mission 9 (40 min),
7) Mission 4 (40 min),	7) Mission 10 (40 min),
8) NASA TLX-II, Trust & Anthropomorphism, and Demographics (30 min)	8) NASA TLX-II, Trust, Anthropomorphism, Demographics, and Debriefing (30 min)
	9) Post-Check Procedure (15 min)

Note. Between two sessions, there were one or two-week intervals. From the hands-on training through the post-check procedure, a 15-minute break was applied after each task; and there was also a half-hour lunch break. Therefore, the total approximate time for the experimental session was eight hours.

The “synthetic agent” pilot communicated and coordinated with the navigator and the photographer in a timely manner but with restricted vocabulary. The vocabulary used by the “synthetic agent” pilot was similar to the vocabulary used by a real-synthetic agent that served as a pilot during a previous experiment [1]. During the training and the task, the navigator and the photographer used templates of allowable utterances provided to them by the researchers to communicate effectively with the synthetic agent via text-chat.

Throughout the experiment, a series of failures or anomalies were injected into the team task that fall within the three categories of: automation failure, autonomy failure, and malicious cyber-attacks. Each failure was applied to selected targets throughout the mission and the teams had a specific time limit to overcome each failure. Whether the team overcame the failure or not, the mission continued. These failures or anomalies are important to the overall

design of the experiment as they have the ability to impact team performance, which is a key metric in this study.

RPAS Degraded Conditions

Automation failure is the first category of RPAS failure which occurs on the pilot and photographer's shared information display. The shared information display contains information such as the current and next waypoint information, distance and time from the current target. In order to overcome the automation failure, teams need to communicate effectively in a timely manner and take a good photo in 400 seconds.

Autonomy failures include two comprehension failures and one anticipation failure which would arise from the synthetic teammate failing to understand a message or to anticipate (such as taking a good photo) the status of the team. In order to overcome the autonomy failures, human team members need to consistently send the correct information about the target waypoint to the synthetic team member

Malicious cyber-attacks are cyber-attacks on the synthetic agent which have the effect of flying the RPA into forbidden enemy territory while the agent denies that this is the destination. The malicious attack happened only once for each team during the last 10-minutes of Mission 10 (i.e., the last mission). If either the navigator or the photographer notices that the RPA is off-route and is going to an enemy designated area, they let intelligence (i.e., the experimenter) know this via a chat message so that the mission will be aborted and the team will have overcome the failure. During training sessions, the experimenters highlighted this specific waypoint to the navigator role and noted that human team members should avoid that waypoint and alert intel that there is a problem if the RPAS is moving towards it.

3.3 Measures

Measures of team performance (mission and target level) and team process (process ratings, communication flow, coordination, situation awareness (i.e., number of failures overcome), and verbal behavior were collected, in addition to measures from the human team members: facial expression, heart rate (ECG), team trust, electrical activity of the brain (EEG), NASA TLX workload [40], and demographics. For this present study, we consider only the following measures:

Team Trust was measured by a modified questionnaire originally developed by Mayer et al. (1995) [32]. In the questionnaire, we asked 25

questions with Likert scale responses ranging from "1" = Strongly Agree to "5" = Strongly disagree. To assess how team trust changed across time, the questionnaire was administered twice: once after each session (i.e., after missions 4 and 10).

Mission level performance score was a weighted composite score containing team level mission parameters, including time spent in warning and alarm states, number of missed targets, and rate of good target photographs per minute (which was weighted the heaviest among the parameters). Each team began with the maximum score of 1000, but then lost points depending on the final values of the mission parameters [41].

Target Processing Efficiency (TPE) score accounted for the time spent inside a target waypoint to get a good photo (higher scores equate to more efficiency, in this case, the maximum score per target is 1000) [41].

Number of failures team overcome. If a team successfully overcame any type of the failure by the end of a mission, then we counted "1" and took the sum across 10 missions. Therefore, we only considered the sum of the failures overcome by each team. The proportion of the 22 teams that overcame failures was approximately equal for automation (65%) and autonomy (64%). However, only 41% of teams overcame malicious cyber-attacks.

4. Results

4.1 High-, Medium-, and Low Performing Teams.

As the relationship of performance and trust is a focal point of this study, we clustered teams into high, medium, and low performing groups. By doing this, it allows for more finite statistical observations to be made across multiple levels of performance.

In order to determine optimal number of team clusters we applied *K-means clustering* analysis on the average mission level team performance score, target performance score, and number of failures overcome to obtain clusters. We chose this analysis technique, because it minimizes the mean measure using Euclidian distance and also seeks to partition the observations into pre-specified number of clusters [42]. In the clustering analysis, we excluded two teams because of the paradoxical relationship between their team score (mission and target level) and number of failures overcome (i.e., high performance but few failures overcome). The analysis was conducted using the "flexclust" 'stats' package [43] in R [44].

During the analysis, first, we randomly initialized two points, called cluster centroids, and then we checked the within cluster sum of squares,

which captures the amount of variability present within all of the clusters. However, we need to find a value of k (i.e., number of cluster) that avoids overfitting the model while clustering the data close to the true empirical distribution. To solve this issue, we chose the “Elbow Method”, which looks at the within groups sum of squares (wss or, equivalently, the percentage of variance explained) as a function of the number of clusters [45]. According to this method, one should choose a number of clusters so that adding another cluster does not give much better modelling of the data. If the wss is obtained for multiple possible values of k , one can plot the wss values and find the point where the marginal drops and an “elbow” is formed (Figure 2). The graph below shows that was sharply drops at 2 clusters. It also drops with a milder slope at 3 clusters. Beyond 3 clusters the graph levels off. Thus, we can suggest that the optimal number of clusters is three ($K = 3$; see Figure 2).

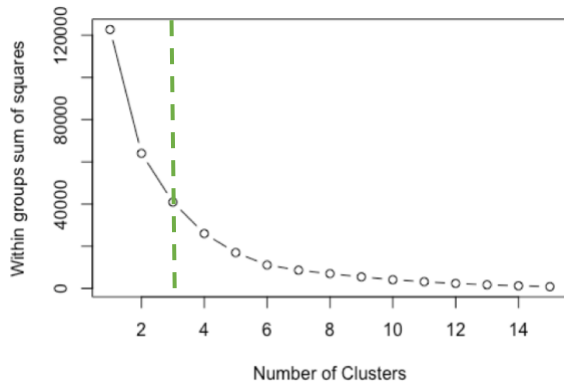


Figure 2. Total within-cluster sum of squares for K-means clustering applied to the average team performance score, target performance scores, and number of failures overcome

4.2 Team Trust

To analyze team trust, we first took the average of the navigator and the photographer responses (Likert scale: “1 = Strongly Disagree” to “5 = Strongly Agree”) for each question. Next, we performed a 3 (condition: Low-, Medium-, and High-Performing Team) \times 2 (session) \times 25 (trust questions) repeated measures Multivariate Analysis of Variance (MANOVA) on 25 questions, for each session, and each performance cluster (condition). Results from a MANOVA are summarized in Table 2.

Table 2. Mixed ANOVA Results

Source	df	F	p	η^2
Question	24	14.458	0.000	0.460

Condition	2	0.324	0.715	0.039
Session	1	4.405	0.051	0.206
Question by Condition	48	1.791	0.002	0.174
Session by Condition	2	0.668	0.526	0.073
Question by Session	24	2.293	0.001	0.119
Question by Session by Condition	48	0.942	0.587	0.100

Based on significant interaction effects, we also conducted pairwise comparisons (based on LSD test) for each interaction, but, due to limited space, we only report significant results from the *independent t-tests*. Turning to the significant *Question \times Condition* effect in the following table, we find that human team members (navigator and photographer) in low performing teams had lower trust with each other than in high performing teams (see Questions 1, 2, 4, 5 and 6 on Table 3). According to Question 3, human team members in medium and high performing teams also trusted the synthetic agent more than in the low performing teams.

Table 3. Mean and Standard Error of Each Significant Question

Source	Low M(SE)	Medium M(SE)	High M(SE)
(1) If I had my way, I would not let nav./ photog. have any influence over issues that are important to me.	3.50* (0.33)	2.88 (0.21)	2.58* (0.27)
(2) I really wish I had a good way to keep an eye on the nav./ photog..	4.00* (0.29)	3.40* (0.18)	2.50* (0.24)
(3) I would tell the synthetic pilot about mistakes I have made on the team task which was critical to me, even if I could not monitor its actions.	3.31* (0.37)	4.35* (0.24)	3.79 (0.31)
(4) I would tell nav./photog. about mistakes I have made on team task, even if they could damage my reputation.	3.50* (0.27)	4.30 (0.17)	4.33* (0.22)
(5) I would share my opinion about sensitive issues with the nav./ photog. even if my opinion were unpopular.	3.25* (0.31)	4.05* (0.19)	3.83 (0.26)
(6) If the nav./ photog. asked why a problem happened, I would speak freely even if I were partly to blame.	3.81* (0.25)	4.60* (0.16)	4.125 (0.21)

Note. Navigator (Nav.) and Photographer (Photo.); * and ** indicate $p < 0.05$ and $p < 0.001$, respectively.

According to the significant interaction effect of *Question \times Sessions*, team members in high performing teams increasingly viewed the

synthetic agent as a machine from Session 1 to Session 2 (i.e., in response to the question, “While chatting with the synthetic pilot, it felt like I was talking to a real person” ($M_{Session1} = 2.76$, $SE_{Session1} = 0.45$, $M_{Session2} = 2.34$, $SE_{Session2} = 0.50$, $p < 0.05$). Also, human team members in high performing teams considered the synthetic agent less enjoyable (i.e., “I enjoyed working with the synthetic agent”, $M_{Session1} = 3.19$, $SE_{Session1} = 0.30$, $Mean_{Session2} = 2.55$, $SE_{Session2} = 0.42$, $p < 0.05$).

Based on the significant session main effect, team trust ($M_{Session1} = 3.24$, $SE_{Session1} = 0.15$, $Mean_{Session2} = 2.35$, $SE_{Session2} = 0.26$, $p < 0.05$) and also trust in the synthetic agent ($M_{Session1} = 4.09$, $SE_{Session1} = 0.20$, $Mean_{Session2} = 3.54$, $SE_{Session1} = 0.25$) significantly decreased from Session 1 to Session 2 ($p < .05$). Probing into the significant main effect of question brought up some interesting results across individual questions. During the task human team members trusted each other more than the synthetic agent (see Table 4).

Table 4. Mean and Standard Deviation of Each Significant Question

Question
(1) if I had my way, I would not let agent ($M = 3.56$, $SE = 0.18$) or human ($M = 2.99$, $SE = 0.16$, $p < .05$) team member have any influence over issues that are important to me.
(2) I would be willing to let agent ($M = 2.30$, $SE = 0.23$) or human team member ($M = 2.91$, $SE = 0.17$, $p < .05$) have complete control over my task in the team.
(3) I really wish I had a good way to keep an eye on the agent ($M = 4.18$, $SE = 0.15$) or human ($M = 3.30$, $SE = 0.14$, $p < .05$) team member.
(4) I would be comfortable giving the agent ($M = 2.67$, $SE = 0.21$) or human ($M = 3.63$, $SE = 0.14$, $p < .05$) team member a task or problem which was critical to me, even if I could not monitor his/her/its actions.
(5) If someone questioned the agent's ($M = 3.21$, $SE = 0.21$) or human team member's ($M = 3.76$, $SE = 0.13$, $p < .05$) motives, I would give the agent/ nav./ photo. the benefit of the doubt.

4.3 Results Overview

The summary of these findings are depicted in Table 5.

Table 5. Mean and Standard Deviation of Each Significant Question, by Performance Cluster (Low, Medium, and High)

Question
(1) Human team members had trust issues with each other: Low = Medium > High
(2) Human team members' trusted the synthetic agent: Medium > Low; High ≥ Low

(3) Team members increasingly viewed the synthetic agent as a machine: Only for **High performing: Session 2 > Session 1**

(4) Trust in synthetic agent: **Session 1 > Session 2**

5. Discussion

This study provides further understanding of the impact an autonomous agent team member can have on trust within a team, that trust can change over time, and low trust causes team performance issues. In general, we can see multiple issues that arise surrounding trust of an autonomous agent team member, particularly in low performing teams.

More specifically, our results indicate that there are lower levels of trust (on multiple aspects of the interpersonal trust survey) in the autonomous agent in low performing teams than both medium and high performing teams. This is significant in that there is clearly a behavior or perception that emerges in low performing teams that results less trust than medium and high performing teams. This highlights the interesting and open question of whether low levels of trust lead to lower performance, or if teams destined to have lower performance also produce lower levels of trust among its team members. In previous studies, the focus has generally been to understand how a single human reacts to autonomy in a team, not how autonomy impacts human-human interaction and behaviors. More research is needed to further investigate how autonomy can impact the team relationship among other human team members.

The second interesting finding from this study is the loss of trust across low, medium, and high performing teams from session 1 to 2. The results indicate that there was more trust in the autonomous agent at the beginning of the teamwork process than at the end of the process. The question becomes why? One potential explanation for this is a compounding effect that human team members were able to think about their experiences with the autonomous agent in session 1 and develop an opinion of the agent (negative based on data) that bled into session 2.

Finally, an interesting result of the interpersonal trust survey is that in addition to the human team members indicating low levels of trust in the autonomous agent, both low and medium performing teams also indicated lower levels of trust in their human team members. For example, the question, *I really wish I had a good way to keep an eye on the mission planner* (human role), was found to be significant in both low and medium performing teams. This indicates that there are issues surrounding human-human team member trust and these issues are mainly found in low and medium

performing teams. High performing teams are not exhibiting the same issues of human-human trust.

This finding also highlights a difference of the high performing teams, compared to the low and medium performing teams. For some reason, high performing teams did not indicate having human-human trust issues (as compared to low and medium). The lack or lessened human-human trust issue may have positively impacted their overall team performance. A deeper exploration is needed to understand these differences and why higher performing teams were able to never develop trust issues or possibly overcome trust issues.

Future studies should investigate if it is the manner in which the autonomous agent interacts with human team members (communication, coordination, cooperation) or if there is a perceptual bias that negatively impacts trust in this context. We also recommend that future studies provide a qualitative component of their data collection allowing human participants to openly express in their own words issues that stem from trust and additional social behavior characteristics. The survey data, in addition to our team performance data, sheds light on trust in a human-autonomy teaming context, but qualitative responses could add an additional layer of depth to our understanding.

Answering these questions is critical to designing more effective human-autonomy teams. It is critical to know if the failures are associated with the actual behavior of the agent or the human's perception. If we know which of these is impacting trust (and it may be both), then we can focus on better developing an autonomous agent that is demonstrating behavior that is viewed as more trustworthy, or if we should implement training protocols for the human to better understand the agent and hopefully improve their perception of the agent.

6. Conclusion

As human-autonomy teams become more prevalent, there is a significant need to better understand the novel factors that can influence their performance. In this study, we aimed to better understand trust in human-autonomy teams, finding that trust is important to team performance. A multitude of interesting questions stem from this study and need to be further studied, such as, 1) is high trust always associated with high team performance?, 2) why do lower performing teams have lower levels of trust in an autonomous agent?, and 3) what causes human team members to lose trust in an autonomous teammate over time? Future studies should focus specifically on the causal direction of the

performance-trust relationship; and how trust is gained or lost in human-autonomy teams, through more granular and qualitative analysis of team process data. Further understanding the development of trust will help to better design human-autonomy teams and inform and build the autonomous agent to account for the concept of trust.

7. Acknowledgments

This research is supported by ONR Award N000141712382 (Program Managers: Marc Steinberg, Micah Clark). The authors acknowledge Steven M. Shope from Sandia Research Corporation who developed the next generation synthetic teammate testbed for the current experiment.

References

- [1] "Our Robots, Ourselves: Robotics and the Myths of Autonomy, David A. Mindell, eBook - Amazon.com." [Online]. Available: <https://www.amazon.com/Our-Robots-Ourselves-Robotics-Autonomy-ebook/dp/B00SI02AWK>. [Accessed: 16-Jun-2018].
- [2] D. D. Woods, "The Risks of Autonomy Doyle's Catch," *Journal of Cognitive Engineering and Decision Making*, vol. 10, no. 2, pp. 131–133, Jun. 2016.
- [3] M. Johnson *et al.*, "Beyond Cooperative Robotics: The Central Role of Interdependence in Coactive Design," *IEEE Intelligent Systems*, vol. 26, no. 3, pp. 81–88, May 2011.
- [4] J. D. Lee, "Perspectives on Automotive Automation and Autonomy," *Journal of Cognitive Engineering and Decision Making*, vol. 12, no. 1, pp. 53–57, Mar. 2018.
- [5] M. R. Endsley, "From Here to Autonomy: Lessons Learned From Human–Automation Research," *Hum Factors*, vol. 59, no. 1, pp. 5–27, Feb. 2017.
- [6] J. M. Bradshaw *et al.*, "Adjustable Autonomy and Human-Agent Teamwork in Practice: An Interim Report on Space Applications," in *Agent Autonomy*, Springer, Boston, MA, 2003, pp. 243–280.
- [7] N. J. McNeese, M. Demir, N. J. Cooke, and C. Myers, "Teaming With a Synthetic Teammate: Insights into Human-Autonomy Teaming," *Hum Factors*, vol. 60, no. 2, pp. 262–273, 2018.
- [8] P. Adler, "Work organization: From Taylorism to teamwork," *Perspectives on Work*, pp. 61–65.
- [9] R. Sun, Ed., *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation*, Y First edition edition. Cambridge; New York: Cambridge University Press, 2005.
- [10] D. Marr, "Artificial intelligence—A personal view," *Artificial Intelligence*, vol. 9, no. 1, pp. 37–48, Aug. 1977.
- [11] J. D. Lee and K. A. See, "Trust in Automation: Designing for Appropriate Reliance," *Hum Factors*, vol. 46, no. 1, pp. 50–80, Mar. 2004.

- [12] B. Muir, "Trust between humans and machines, and the design of decision aids - ScienceDirect," *Journal of Man-Machine Studies*, vol. 27, no. 5-6, pp. 527-539, 1987.
- [13] K. A. Hoff and M. Bashir, "Trust in Automation Integrating Empirical Evidence on Factors That Influence Trust," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 57, no. 3, pp. 407-434, May 2015.
- [14] K. E. Schaefer, J. Y. C. Chen, J. L. Szalma, and P. A. Hancock, "A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems," *Hum Factors*, vol. 58, no. 3, pp. 377-400, May 2016.
- [15] J. Y. C. Chen and M. J. Barnes, "Human-Agent Teaming for Multirobot Control: A Review of Human Factors Issues," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 1, pp. 13-29, Feb. 2014.
- [16] M. Yuan Zhang and X. Jessie Yang, "Evaluating effects of workload on trust in automation, attention allocation and dual-task performance," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, no. 1, pp. 1799-1803, Sep. 2017.
- [17] R. S. Gutzwiller, S. H. Espinosa, C. Kenny, and D. S. Lange, "A Design Pattern for Working Agreements in Human-Autonomy Teaming," in *Advances in Human Factors in Simulation and Modeling*, Springer, Cham, 2017, pp. 12-24.
- [18] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 30, no. 3, pp. 286-297, May 2000.
- [19] A. van Wissen, Y. Gal, B. A. Kamphorst, and M. V. Dignum, "Human-agent teamwork in dynamic environments," *Computers in Human Behavior*, vol. 28, no. 1, pp. 23-33, Jan. 2012.
- [20] J. M. Bradshaw, V. Dignum, C. Jonker, and M. Sierhuis, "Human-agent-robot teamwork," *IEEE Intelligent Systems*, vol. 27, no. 2, pp. 8-13, Mar. 2012.
- [21] G. Klein, D. D. Woods, J. M. Bradshaw, R. R. Hoffman, and P. J. Feltovich, "Ten challenges for making automation a 'team player' in joint human-agent activity," *IEEE Intelligent Systems*, vol. 19, no. 6, pp. 91-95, Nov. 2004.
- [22] N. J. Cooke, J. C. Gorman, C. W. Myers, and J. L. Duran, "Interactive Team Cognition," *Cognitive Science*, vol. 37, no. 2, pp. 255-285, Mar. 2013.
- [23] J. Ball *et al.*, "The synthetic teammate project," *Comput Math Organ Theory*, vol. 16, no. 3, pp. 271-299, Sep. 2010.
- [24] M. Demir, N. J. McNeese, and N. J. Cooke, "Team situation awareness within the context of human-autonomy teaming," *Cognitive Systems Research*, vol. 46, pp. 3-12, Dec. 2017.
- [25] M. Demir, N. J. McNeese, and N. J. Cooke, "Team Synchrony in Human-Autonomy Teaming," in *Advances in Human Factors in Robots and Unmanned Systems*, vol. 595, J. Chen, Ed. Cham: Springer International Publishing, 2017, pp. 303-312.
- [26] X. Fan, M. McNeese, and J. Yen, "NDM-Based Cognitive Agents for Supporting Decision-Making Teams," *Human-Computer Interaction*, vol. 25, no. 3, pp. 195-234, Aug. 2010.
- [27] J. C. Walliser, P. R. Mead, and T. H. Shaw, "The Perception of Teamwork With an Autonomous Agent Enhances Affect and Performance Outcomes," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, no. 1, pp. 231-235, Sep. 2017.
- [28] E. K. Chiou and J. D. Lee, "Cooperation in Human-Agent Systems to Support Resilience: A Microworld Experiment," *Hum Factors*, vol. 58, no. 6, pp. 846-863, Sep. 2016.
- [29] E. K. Chiou, J. Lee, and T. Su, "Reciprocal exchange in human-agent cooperation," *Computers in Human Behavior*, In Press.
- [30] D. J. McAllister, "Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations," *ACAD MANAGE J*, vol. 38, no. 1, pp. 24-59, Feb. 1995.
- [31] K. Sycara and G. Sukthankar, "Literature Review of Teamwork Models," Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, Tech Report CMU-RI-TR-06-50, Nov. 2006.
- [32] R. C. Mayer, J. H. Davis, and F. D. Schoorman, "An Integrative Model of Organizational Trust," *The Academy of Management Review*, vol. 20, no. 3, pp. 709-734, 1995.
- [33] J. M. Bradshaw, "Making Agents Acceptable to People," in *Multi-Agent Systems and Applications III*, Springer, Berlin, Heidelberg, 2003, pp. 1-3.
- [34] K. E. Oleson, D. R. Billings, V. Kocsis, J. Y. C. Chen, and P. A. Hancock, "Antecedents of trust in human-robot collaborations," in *2011 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2011, pp. 175-178.
- [35] E. J. de Visser *et al.*, "Almost human: Anthropomorphism increases trust resilience in cognitive agents," *Journal of Experimental Psychology: Applied*, vol. 22, no. 3, pp. 331-349, Sep. 2016.
- [36] J. E. Mercado, M. A. Rupp, J. Y. C. Chen, M. J. Barnes, D. Barber, and K. Procci, "Intelligent Agent Transparency in Human-Agent Teaming for Multi-UxV Management," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 58, no. 3, pp. 401-415, May 2016.
- [37] K. E. Schaefer, E. R. Straub, J. Y. C. Chen, J. Putney, and A. W. Evans, "Communicating intent to develop shared situation awareness and engender trust in human-agent teams," *Cognitive Systems Research*, vol. 46, pp. 26-39, Dec. 2017.
- [38] E. J. de Visser, R. Pak, and T. H. Shaw, "From 'automation' to 'autonomy': the importance of trust

- repair in human-machine interaction,” *Ergonomics*, vol. 0, no. 0, pp. 1–19, Mar. 2018.
- [39] N. J. Cooke and S. M. Shope, “Designing a Synthetic Task Environment,” in *Scaled Worlds: Development, Validation, and Application*, L. R. E. Schiflett, E. Salas, and M. D. Covert, Eds. Surrey, England: Ashgate Publishing, 2004, pp. 263–278.
 - [40] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research,” in *Human Mental Workload*, P. A. Hancock and N. Mashkati, Eds. Amsterdam: North Holland Press, 1988.
 - [41] N. J. Cooke *et al.*, “Acquisition and Retention of Team Coordination in Command-and-Control,” Jul. 2007.
 - [42] G. James, D. Witten, T. Hastie, and R. Tibshirani, Eds., *An introduction to statistical learning: with applications in R*. New York: Springer, 2013.
 - [43] F. Leisch and E. Dimitriadou, *flexclust: Flexible Cluster Algorithms*. 2018.
 - [44] R Development Core Team, “R: A Programming Environment for Data Analysis and Graphics (Version 3.3.2),” 31-Oct-2016. [Online]. Available: <https://www.r-project.org/>. [Accessed: 24-Jan-2017].
 - [45] M. Sarstedt and E. Mooi, “Cluster Analysis,” in *A Concise Guide to Market Research*, Springer, Berlin, Heidelberg, 2014, pp. 273–324.
 - [46] Jones, Gareth R., and Jennifer M. George. "The experience and evolution of trust: Implications for cooperation and teamwork." *Academy of management review* 23.3 (1998): 531-546.
 - [47] Sheng, Chieh-Wen, Yi-Fang Tian, and Ming-Chia Chen. "Relationships among teamwork behavior, trust, perceived team support, and team commitment." *Social Behavior and Personality: an international journal* 38, no. 10 (2010): 1297-1305.
 - [48] Erdem, Ferda, and Janset Ozen. "Cognitive and affective dimensions of trust in developing team performance." *Team Performance Management: An International Journal* 9.5/6 (2003): 131-135.